



GENOME ANNOTATION AND TRANSCRIPTOMICS OF OIL-PRODUCING ALGAE

Sabeeha Merchant
UNIVERSITY OF CALIFORNIA LOS ANGELES

03/19/2015
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTE
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE					<i>Form Approved</i> OMB No. 0704-0188							
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.</p>												
1. REPORT DATE (DD-MM-YYYY) 16-03-2015		2. REPORT TYPE FINAL PERFORMANCE REPORT			3. DATES COVERED (From - To) From 04-01-2010 To 12-31-2014							
4. TITLE AND SUBTITLE GENOME ANNOTATION AND TRANSCRIPTOMICS OF OIL-PRODUCING ALGAE					5a. CONTRACT NUMBER FA9550-10-1-0095							
					5b. GRANT NUMBER							
					5c. PROGRAM ELEMENT NUMBER							
6. AUTHOR(S) Merchant, Sabeeha Pellegrini, Matteo					5d. PROJECT NUMBER							
					5e. TASK NUMBER							
					5f. WORK UNIT NUMBER							
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) UNIVERSITY OF CALIFORNIA, LOS ANGELES OFFICE OF RESEARCH ADMINISTRATION 11000 KINROSS AVE STE 102 LOS ANGELES, CA 90095-2000					8. PERFORMING ORGANIZATION REPORT NUMBER UCLA PATS Number 20093681							
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) USAF, AFRL DUNS 143574726 AF OFFICE OF SCIENTIFIC RESEARCH 875 NORTH RANDOLPH STREET, RM 3112 ARLINGTON VA 22203					10. SPONSOR/MONITOR'S ACRONYM(S) AFOSR/PKR1							
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)							
12. DISTRIBUTION/AVAILABILITY STATEMENT public availability												
13. SUPPLEMENTARY NOTES												
14. ABSTRACT Most algae accumulate triacylglycerols (TAGs) when they are starved for essential nutrients like N, S, P (or Si in the case of some diatoms). We had proposed to use whole transcriptome analyses to detail the changes in gene expression that occur during N-starvation induced TAG accumulation in Chlamydomonas. We used RNA-Seq on the Illumina platform for quantitative determination of the Chlamydomonas transcriptome. Deep coverage over multiple time points in nearly a dozen different experiments in at least 3 strains allowed us to distinguish early responses to the stress signal and to unequivocally identify relevant changes in RNA abundance. This led to the functional identification of 8 acyltransferases, several of which were documented to have a role in TAG accumulation in stressed algal cells, and candidate regulators of the N starvation response. We used our RNA-Seq data analysis pipeline for building and quantitating transcripts from Chlorella and Cyclotella in TAG-producing conditions and we assembled a draft genome for Cyclotella. During a period of project extension, we generated new strand-specific RNA-Seq data from ribosomal RNA-depleted preparations of RNA for the purpose of identifying regulatory lncRNAs.												
15. SUBJECT TERMS RNA-Seq, Chlamydomonas, biodiesel, triacylglycerol, Cyclotella, transcriptomics, regulation, stress												
16. SECURITY CLASSIFICATION OF: <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%; padding: 2px;">a. REPORT</td> <td style="width: 33%; padding: 2px;">b. ABSTRACT</td> <td style="width: 33%; padding: 2px;">c. THIS PAGE</td> </tr> <tr> <td style="text-align: center; padding: 2px;">UU</td> <td style="text-align: center; padding: 2px;">UU</td> <td style="text-align: center; padding: 2px;">UU</td> </tr> </table>			a. REPORT	b. ABSTRACT	c. THIS PAGE	UU	UU	UU	17. LIMITATION OF ABSTRACT UU		18. NUMBER OF PAGES 4	
a. REPORT	b. ABSTRACT	c. THIS PAGE										
UU	UU	UU										
19a. NAME OF RESPONSIBLE PERSON Sabeeha Merchant					19b. TELEPHONE NUMBER (Include area code) 310-825-8300							

Reset

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Abstract

Most algae accumulate triacylglycerols (TAGs) when they are starved for essential nutrients like N, S, P (or Si in the case of some diatoms). We had proposed to use whole transcriptome analyses to detail the changes in gene expression that occur during N-starvation induced TAG accumulation in *Chlamydomonas*. We used RNA-Seq on the Illumina platform for quantitative determination of the *Chlamydomonas* transcriptome. Deep coverage over multiple time points in nearly a dozen different experiments in at least 3 strains allowed us to distinguish early responses to the stress signal and to unequivocally identify relevant changes in RNA abundance. This led to the functional identification of 8 acyltransferases, several of which were documented to have a role in TAG accumulation in stressed algal cells, and candidate regulators of the N starvation response. We used our RNA-Seq data analysis pipeline for building and quantitating transcripts from *Chlorella* and *Cyclotella* in TAG-producing conditions and we assembled a draft genome for *Cyclotella*. During a period of project extension, we generated new strand-specific RNA-Seq data from ribosomal RNA-depleted preparations of RNA for the purpose of identifying regulatory lncRNAs. Some of these candidates, which map to intergenic regions, show sharp transient patterns of expression during the cell cycle consistent with a regulatory role.

Key Accomplishments

1. The analysis of the *Chlamydomonas* transcriptome in N-starved cells in 3 time course experiments is complete. We have curated enzymes involved in TAG metabolism. Besides the 5 previously described *DGTT1-DGTT5* genes (encoding type 2 diacylglycerol acyltransferases), we identified two additional genes, *DGAT1* and *DGAT3*, encoding distinct candidate enzymes. *DGAT1* had eluded prior discovery because of incomplete sequence coverage in that region of the genome. By using new sequence data and manual assembly, we have increased the sequence coverage of *DGAT1* from ~ 50% to ~ 90%. *DGTT1*, *DGAT1* and *DGAT3* are coordinately expressed in N-starved *Chlamydomonas* cells. The genes are also up-regulated in other stress situations that promote TAG accumulation, consistent with a causal connection. We developed an in vivo method for testing the function of algal genes in TAG synthesis. *DGAT3*, conserved in the plant lineage, does not support TAG synthesis in this assay, and we wonder whether it may have a regulatory role.
2. We have developed methodology for and now have considerable experience with the analysis of Illumina sequence data for transcriptome studies, and also for genome and transcript assembly. These methods have been applied to transcriptome data from *Chlorella* (sub-contract to Sayre) and *Cyclotella* (sub-contract to Hildebrand). In both cases the reads were used to assemble better transcript models and estimate transcript abundance under specific situations.
3. We also expanded the scope of the project (without increases in cost) to include genome sequencing of *Cyclotella cryptica* in collaboration with the Hildebrand group. We generated libraries from DNA and RNA and performed *de novo* assembly to generate a first draft of the genome and transcriptome of *Cyclotella cryptica*. The first draft of the genome has an N50 (i.e. typical fragment size) of 10kb and a total genome size of approximately 160Mb. The total genome estimate is in agreement with that generated previously based on calorimetric measurements. The transcript-based gene models include alternative forms generated by splicing or from different start sites. Gene models were constructed by using MAKER, which combines multiple gene prediction tool including Augustus and Fgenesh. We found that Augustus models that overlapped Maker models were the most reliable, and allowed us to generate a high confidence set of approximately 10,000 genes. We also generated bisulfite data to produce a map of cytosine methylation, and found that this mark is associated with repeat regions of the genome. The manuscript describing these results is in preparation.

4. We generated new strand-specific RNA-Seq data from rRNA-depleted samples with parallel ChIP-Seq data from 18 samples collected during the *Chlamydomonas* cell cycle. We used these reads to assemble nearly 900 new transcripts that appear to be long non-coding RNAs (lncRNAs). These lncRNAs map to intergenic regions and also to the opposite strand of protein-coding mRNAs. In individual cases that were manually curated, the RNAs do not appear to code for long proteins and their pattern of expression is transient, suggesting that they may be regulatory. Manual curation also indicates that the 5' ends have chromatin marks that are typical for PolIII transcripts, consistent with these being lncRNAs. We are very interested in continuing this work should funding permit.

Background

Most algae accumulate triacylglycerols (TAGs) when they are starved for essential nutrients like N, S, P (or Si in the case of some diatoms). In the absence of such essential nutrients they are unable to synthesize macromolecules that require these elements. Therefore, they cannot grow; rather, they divert carbon towards storage molecules – either starch or neutral lipids like TAGs. The TAGs are precursors for biodiesel because the fatty acid constituents of TAG can be transesterified to generate methyl esters.

We had proposed to use transcriptome approaches to detail the changes in gene expression that occur during N-starvation induced TAG accumulation in *Chlamydomonas*. *Chlamydomonas* is a key reference organism for the chlorophyte algae because there is a draft genome with over 17,000 gene models (many of them manually curated and functionally annotated) and there are resources for classical and reverse genetics.

We had also proposed to undertake RNA Seq analysis of *Chlorella* (for another grantee Richard Sayre at the Donald Danforth Plant Science Center at the time of the award) and to assemble a transcriptome and genome for a diatom *Cyclotella cryptica* (for another grantee Mark Hildebrand at Scripps / UCSD). *Chlorella* was of interest because it grows to high density and the diatom had been noted previously to be a high producer of TAGs. *Cyclotella* is a promising alga for biofuel production, but to date virtually nothing is known about its genome and genes.

Approach

We used RNA-Seq on the Illumina platform for quantitative determination of the *Chlamydomonas* transcriptome. Deep coverage over multiple time points in nearly a dozen different experiments in at least 3 strains allowed us to distinguish early responses to the stress signal and to unequivocally identify relevant changes in RNA abundance. Gene models were corrected based on RNA Seq coverage, which allowed the proteins to be correctly expressed in heterologous systems to validate the activity of the gene product. We also used a reverse-genetic strategy as a form of validation of the role of individual genes in the TAG accumulation pathway. The success on this project prompted us to use the RNA-Seq coverage to annotate new (previously-unobserved) transcripts from the *Chlamydomonas* genome, many of which correspond to long non-coding RNAs (lncRNAs).

Significance

The fact that multiple distinct acyltransferase enzymes are upregulated during N and other stress starvation in *Chlamydomonas* indicates that there may be several sites / foci of TAG synthesis.

The transient pattern of expression of lncRNAs suggests that they may be regulatory. Hence we are interested in pursuing this work to understand how lncRNAs impact the metabolic program, especially during the transition from carbon fixing to carbon utilizing stage of the cell cycle.

Collaborators

Christoph Benning (Michigan State University) for chemical analysis of fatty acids in TAG and TAG quantitation

Arthur Grossman (Carnegie Institution) for screening for loss of function mutations in various genes.

Mark Hildebrand (UCSD/SIO) for biology of *Cyclotella*

Key findings

1. RNA-seq coverage-based annotation of gene models generates more reliable and functional gene models.
2. We identified several TAG synthesis enzymes that had not previously been described in *Chlamydomonas* and documented their activities by functional complementation in yeast.
3. We showed that TAG accumulation occurred in several different stress situations that impact cell growth and division, with evidence for increased expression of the diacylglycerol acyltransferase-encoding genes in each case.

Recognition of the investigators and the work

Post-doctoral researcher Blaby has received a post-doctoral award from the Molecular Biology Institute at UCLA.

Former post-doctoral researcher Boyle is now an independent investigator at Colorado School of Mines.

Merchant has been honored by election to the American Academy of Arts and Sciences and the US National Academy of Science. She has also been appointed as an external member of the Max Planck Society.

The publications from this project are very well-cited (> 240 to date).

Archival publications

Ferris, P., Olson, B.J., de Hoff, P.L., Douglass, S., Casero, D., Prochnik, S., Geng, S., Rai, R., Grimwood, J., Schmutz, J., Nishii, I., Hamaji, T., Nozaki, H., Pellegrini, M., Umen, J.G. (2010) Evolution of an expanded sex-determining locus in *Volvox*. *Science* 328: 351-354.

<http://www.sciencemag.org/content/328/5976/351.long>

cited 51 times

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Pellegrini, M., Merchant, S.S., Malasarn, D. (2011) A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *Plant J.* 66:770-80.

<http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2011.04537.x/full>

cited 39 times

Boyle, N.R., Page, M.D., Liu, B., Blaby, I.K., Casero, D., Kropat, J., Cokus, S.J., Hong-Hermesdorf, A., Shaw, J., Karpowicz, S.J., Gallaher, S.D., Johnson, S., Benning, C., Pellegrini, M., Grossman, A.R., Merchant, S.S. (2012) Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J. Biol. Chem.* 287: 15811-25.

<http://www.ncbi.nlm.nih.gov/pubmed/22403401>

cited 70 times

Merchant, S.S., Kropat, J., Liu, B., Shaw, J., Warakanont, J. (2012) TAG, You're It! *Chlamydomonas* as a reference organism for understanding algal triacylglycerol accumulation. *Curr. Opin. Biotechnol.* 23:352-363.

<http://www.sciencedirect.com/science/article/pii/S0958166911007440>

cited 64 times

Blaby, I.K., Casero, D., Glaesener, A.G., Fitz-Gibbon, S., Mettler, T., Gallaher, S., Liu, B., Boyle, N.R., Kropat, J., Stitt, M., Johnson, S., Benning, C., Pellegrini, M., Merchant, S.S. (2013) Systems-level analysis of N-starvation induced TAG accumulation in a *Chlamydomonas* starchless mutant. *Plant Cell* 25: 4305-4323.

<http://www.plantcell.org/content/25/11/4305.full.pdf+html>

cited 15 times

Urzica, E.I., Vieler, A., Hong-Hermesdorf, A., Page, M.D., Casero, D., Gallaher, S.D., Kropat, J., Pellegrini, M., Benning, C., Merchant, S.S. (2013) Remodeling of membrane lipids in iron starved Chlamydomonas. **J. Biol. Chem.** 288:30246-30258.

<http://www.jbc.org/content/288/42/30246.long>

cited 4 times

1.

1. Report Type

Final Report

Primary Contact E-mail

Contact email if there is a problem with the report.

merchant@chem.ucla.edu

Primary Contact Phone Number

Contact phone number if there is a problem with the report

3108258300

Organization / Institution name

Regents of the University of California

Grant/Contract Title

The full title of the funded effort.

Genome annotation and transcriptomics of oil-producing algae

Grant/Contract Number

AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-10-1-0095

Principal Investigator Name

The full name of the principal investigator on the grant or contract.

sabeeha merchant

Program Manager

The AFOSR Program Manager currently assigned to the award

Patrick Bradshaw

Reporting Period Start Date

04/01/2010

Reporting Period End Date

12/31/2014

Abstract

Most algae accumulate triacylglycerols (TAGs) when they are starved for essential nutrients like N, S, P (or Si in the case of some diatoms). We had proposed to use whole transcriptome analyses to detail the changes in gene expression that occur during N-starvation induced TAG accumulation in *Chlamydomonas*. We used RNA-Seq on the Illumina platform for quantitative determination of the *Chlamydomonas* transcriptome. Deep coverage over multiple time points in nearly a dozen different experiments in at least 3 strains allowed us to distinguish early responses to the stress signal and to unequivocally identify relevant changes in RNA abundance. This led to the functional identification of 8 acyltransferases, several of which were documented to have a role in TAG accumulation in stressed algal cells, and candidate regulators of the N starvation response. We used our RNA-Seq data analysis pipeline for building and quantitating transcripts from *Chlorella* and *Cyclotella* in TAG-producing conditions and we assembled a draft genome for *Cyclotella*. During a period of project extension, we generated new strand-specific RNA-Seq data from ribosomal RNA-depleted preparations of RNA for the purpose of identifying regulatory lncRNAs. Some of these candidates, which map to

intergenic regions, show sharp transient patterns of expression during the cell cycle consistent with a regulatory role.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found [here](#). Please do not password protect or secure the PDF. The maximum file size for an SF298 is 50MB.

[AFD-070820-035 merchant 20150316.pdf](#)

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF. The maximum file size for the Report Document is 50MB.

[merchant final report.pdf](#)

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Ferris, P., Olson, B.J., de Hoff, P.L., Douglass, S., Casero, D., Prochnik, S., Geng, S., Rai, R., Grimwood, J., Schmutz, J., Nishii, I., Hamaji, T., Nozaki, H., Pellegrini, M., Umen, J.G. (2010) Evolution of an expanded sex-determining locus in *Volvox*. *Science* 328: 351-354.
<http://www.sciencemag.org/content/328/5976/351.long>
cited 51 times

Kropat, J., Hong-Hermesdorf, A., Casero, D., Ent, P., Pellegrini, M., Merchant, S.S., Malasarn, D. (2011) A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*. *Plant J.* 66:770-80.
<http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2011.04537.x/full>
cited 39 times

Boyle, N.R., Page, M.D., Liu, B., Blaby, I.K., Casero, D., Kropat, J., Cokus, S.J., Hong-Hermesdorf, A., Shaw, J., Karpowicz, S.J., Gallaher, S.D., Johnson, S., Benning, C., Pellegrini, M., Grossman, A.R., Merchant, S.S. (2012) Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J. Biol. Chem.* 287: 15811-25.
<http://www.ncbi.nlm.nih.gov/pubmed/22403401>
cited 70 times

Merchant, S.S., Kropat, J., Liu, B., Shaw, J., Warakanont, J. (2012) TAG, You're It! *Chlamydomonas* as a reference organism for understanding algal triacylglycerol accumulation. *Curr. Opin. Biotechnol.* 23:352-363.
<http://www.sciencedirect.com/science/article/pii/S0958166911007440>
cited 64 times

Blaby, I.K., Casero, D., Glaesener, A.G., Fitz-Gibbon, S., Mettler, T., Gallaher, S., Liu, B., Boyle, N.R., Kropat, J., Stitt, M., Johnson, S., Benning, C., Pellegrini, M., Merchant, S.S. (2013) Systems-level analysis of N-starvation induced TAG accumulation in a *Chlamydomonas* starchless mutant. *Plant Cell* 25: 4305-4323.
<http://www.plantcell.org/content/25/11/4305.full.pdf+html>
cited 15 times

Urzica, E.I., Vieler, A., Hong-Hermesdorf, A., Page, M.D., Casero, D., Gallaher, S.D., Kropat, J., Pellegrini, M., Benning, C., Merchant, S.S. (2013) Remodeling of membrane lipids in iron starved Chlamydomonas. J. Biol. Chem. 288:30246-30258.
<http://www.jbc.org/content/288/42/30246.long>
cited 4 times

Changes in research objectives (if any):

Not applicable

Change in AFOSR Program Manager, if any:

Dr. Kozumbo was replaced by Dr. Bradshaw during the course of the project

Extensions granted or milestones slipped, if any:

We received a 2 year extension to initiate work on lncRNAs.

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, \$K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Mar 17, 2015 16:41:48 Success: Email Sent to: merchant@chem.ucla.edu